

New rigorous perturbation bounds for the generalized Cholesky factorization*

Hanyu Li,[†] Yanfei Yang

College of Mathematics and Statistics, Chongqing University, Chongqing, 401331, P. R. China

Abstract

Some new rigorous perturbation bounds for the generalized Cholesky factorization with normwise or componentwise perturbations in the given matrix are obtained, where the componentwise perturbation has the form of backward rounding error for the generalized Cholesky factorization algorithm. These bounds can be much tighter than some existing ones while the conditions for them to hold are simple and moderate.

AMS classification: 15A23; 15A45

Keywords: Generalized Cholesky factorization; Rigorous perturbation bound; Normwise perturbation; Componentwise perturbation

1 Introduction

Let $\mathbb{R}^{m \times n}$ be the set of $m \times n$ real matrices and $\mathbb{R}_r^{m \times n}$ be the subset of $\mathbb{R}^{m \times n}$ consisting of matrices with rank r . Let I_r be the identity matrix of order r and A^T be the transpose of the matrix A .

Consider the following block matrix

$$K = \begin{bmatrix} A & B^T \\ B & -C \end{bmatrix} \in \mathbb{R}^{(m+n) \times (m+n)}, \quad (1.1)$$

where $A \in \mathbb{R}_m^{m \times m}$ is symmetric positive definite, $B \in \mathbb{R}_n^{n \times m}$, and $C \in \mathbb{R}_n^{n \times n}$ is symmetric positive semi-definite. For this matrix, there always exists the following factorization

$$K = LJ_{m+n}L^T, \quad (1.2)$$

where

$$L = \begin{bmatrix} L_{11} & 0 \\ L_{21} & L_{22} \end{bmatrix}, \quad J_{m+n} = \begin{bmatrix} I_m & 0 \\ 0 & -I_n \end{bmatrix},$$

$L_{11} \in \mathbb{R}_m^{m \times m}$ and $L_{22} \in \mathbb{R}_n^{n \times n}$ are lower triangular, and $L_{21} \in \mathbb{R}_n^{n \times m}$. The factorization (1.2) is called the generalized Cholesky factorization and L is referred to as the generalized Cholesky factor [1]. If the diagonal elements of the lower triangular matrices L_{11} and L_{22} are positive, the factorization is unique.

For the generalized Cholesky factorization, some scholars considered its applications, algorithms, algorithms' numerical stability, and perturbation analysis [1–6]. Several first-order perturbation bounds were presented [2, 4–6]. Since, in some cases, it is unclear whether the first-order bound is a good approximate bound as it ignores the higher-order terms, we have to be careful to use them in practice.

*The work is supported by the National Natural Science Foundation of China (Grant Nos. 11201507; 11171361).

[†]Corresponding author. Email addresses: lihy.hy@gmail.com or hyl@cq.cqu.edu.cn; yangyanfei2008@yeah.net

Oppositely, the rigorous perturbation bounds can be used safely for all cases. So it is important to derive the rigorous bounds for the generalized Cholesky factorization. At present, some rigorous bounds have been given for this factorization [3, 4, 6]. However, these bounds are either quite loose or derived under more restrictive conditions or expensive to compute. The rigorous bounds derived in this paper, using the combination of the classic and refined matrix equation approaches [7], overcome these disadvantages to some extent. They can be much tighter than some existing bounds while the conditions for them to hold are simple and moderate.

The rest of this paper is organized as follows. Section 2 presents some notation and basics. The rigorous perturbation bounds with normwise or componentwise perturbations are given in Sections 3 and 4, respectively. Finally, the concluding remarks of the whole paper is provided.

2 Notation and basics

Given a matrix $A \in \mathbb{R}_r^{m \times n}$, $\|A\|_2$ and $\|A\|_F$ stand for its spectral norm and Frobenius norm, respectively. From [8, pp. 80], we have

$$\|XYZ\|_2 \leq \|X\|_2 \|Y\|_2 \|Z\|_2, \quad \|XYZ\|_F \leq \|X\|_2 \|Y\|_F \|Z\|_2, \quad (2.1)$$

whenever the matrix product XYZ is defined. If A is nonsingular, we denote its standard condition number by $\kappa_2(A) = \|A^{-1}\|_2 \|A\|_2$ and Bauer–Skeel condition number by $\text{cond}_F(A) = \| |A^{-1}| |A| \|_F$ [8, pp. 128]. Here, for any matrix $X = (x_{ij})$, $|X|$ is defined by $|X| = (|x_{ij}|)$.

For any matrix $A = (a_{ij}) \in \mathbb{R}^{n \times n}$, define

$$\text{up}(A) = \begin{bmatrix} \frac{1}{2}a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & \frac{1}{2}a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{2}a_{nn} \end{bmatrix}.$$

Obviously,

$$\|\text{up}(A)\|_F \leq \|A\|_F. \quad (2.2)$$

If $A^T = A$, from [9], we have

$$\|\text{up}(A)\|_F \leq \frac{1}{\sqrt{2}} \|A\|_F. \quad (2.3)$$

Moreover, let $\mathbb{D}_n \in \mathbb{R}^{n \times n}$ be the set of $n \times n$ positive definite diagonal matrices. Then, for any $D_n \in \mathbb{D}_n$,

$$\text{up}(AD_n) = \text{up}(A)D_n. \quad (2.4)$$

The following lemma is needed later in this paper, which is taken from [7].

Lemma 2.1. *Let $a, b > 0$, and $c(\cdot)$ be a continuous function of a parameter $t \in [0, 1]$ such that $b^2 - 4ac(t) > 0$ holds for all t . Suppose that a continuous function $x(t)$ satisfies the quadratic inequality $ax^2(t) - bx(t) + c(t) \geq 0$. If $c(0) = x(0) = 0$, then*

$$x(1) \leq \frac{1}{2a} \left(b - \sqrt{b^2 - 4ac(1)} \right).$$

3 Rigorous perturbation bounds with normwise perturbation

The main theorem, similar to Theorem 3.1 of [7], which is concerned with the regular Cholesky factorization, is presented as follows.

Theorem 3.1. *Let $K \in \mathbb{R}^{(m+n) \times (m+n)}$ be expressed as in (1.1) and factorized as in (1.2). Let $\Delta K \in \mathbb{R}^{(m+n) \times (m+n)}$ be symmetric. If*

$$\|L^{-1}\|_2^2 \|\Delta K\|_F < \frac{1}{2}, \quad (3.1)$$

then $K + \Delta K$ has the following generalized Cholesky factorization

$$K + \Delta K = (L + \Delta L) J_{m+n} (L + \Delta L)^T. \quad (3.2)$$

Moreover,

$$\|\Delta L\|_F \leq \frac{\sqrt{2} \|L^{-1}\|_2 \left[\inf_{D_{m+n} \in \mathbb{D}_{m+n}} \kappa_2(LD_{m+n}^{-1}) \right]}{\sqrt{2} - 1 + \sqrt{1 - 2 \|L^{-1}\|_2^2 \|\Delta K\|_F}} \|\Delta K\|_F \quad (3.3)$$

$$\leq (2 + \sqrt{2}) \|L^{-1}\|_2 \left[\inf_{D_{m+n} \in \mathbb{D}_{m+n}} \kappa_2(LD_{m+n}^{-1}) \right] \|\Delta K\|_F. \quad (3.4)$$

Proof. Using (2.1) and noting the condition (3.1), we have that for any $t \in [0, 1]$,

$$\rho(L^{-1}t(\Delta K)L^{-T}) \leq \|L^{-1}t(\Delta K)L^{-T}\|_F \leq \|L^{-1}\|_2^2 \|\Delta K\|_F < \frac{1}{2}.$$

Here, for any square matrix X , $\rho(X)$ denotes its spectral radius. Thus, from Theorem 2.2 in [2] and its proof or Theorem 2.1 in [6] and its proof, it follows that the matrix $K + t(\Delta K)$ has the following generalized Cholesky factorization

$$K + t(\Delta K) = L(t) J_{m+n} L^T(t) = (L + (\Delta L(t))) J_{m+n} (L + (\Delta L(t)))^T, \quad (3.5)$$

where $L(t)$ is lower triangular having the same structure as that of L in (1.2) and $\Delta L(t) = L(t) - L$ with $\Delta L(0) = 0$. Setting $\Delta L(1) = \Delta L$ in (3.5) gives (3.2).

In the following, we consider (3.3) and (3.4). Observing (1.2), it follows from (3.5) that

$$t(\Delta K) = L J_{m+n} (\Delta L(t))^T + (\Delta L(t)) J_{m+n} L^T + (\Delta L(t)) J_{m+n} (\Delta L(t))^T.$$

Left-multiplying the above equation by L^{-1} and right-multiplying it by L^{-T} gives

$$J_{m+n} (\Delta L(t))^T L^{-T} + L^{-1} (\Delta L(t)) J_{m+n} = t L^{-1} (\Delta K) L^{-T} - L^{-1} (\Delta L(t)) J_{m+n} (\Delta L(t))^T L^{-T}. \quad (3.6)$$

Since $J_{m+n} (\Delta L(t))^T L^{-T}$ is upper triangular, using the symbol "up," from (3.6), we have

$$J_{m+n} (\Delta L(t))^T L^{-T} = \text{up} [t L^{-1} (\Delta K) L^{-T} - L^{-1} (\Delta L(t)) J_{m+n} (\Delta L(t))^T L^{-T}]. \quad (3.7)$$

Taking the Frobenius norm on (3.7) and considering (2.3) and (2.1) leads to

$$\begin{aligned} \|L^{-1} (\Delta L(t))\|_F &= \|J_{m+n} (\Delta L(t))^T L^{-T}\|_F \leq \frac{1}{\sqrt{2}} \|t L^{-1} (\Delta K) L^{-T} - L^{-1} (\Delta L(t)) J_{m+n} (\Delta L(t))^T L^{-T}\|_F \\ &\leq \frac{1}{\sqrt{2}} \left(\|L^{-1}\|_2^2 \|\Delta K\|_F t + \|L^{-1} (\Delta L(t))\|_F^2 \right). \end{aligned}$$

Let $x(t) = \|L^{-1}(\Delta L(t))\|_F$ and $c(t) = \|L^{-1}\|_2^2 \|\Delta K\|_F t$. It is easy to find that both $x(t)$ and $c(t)$ are continuous with respect to t . Moreover,

$$x^2(t) - \sqrt{2}x(t) + c(t) \geq 0.$$

From (3.1), it is seen that for any $t \in [0, 1]$, $2 - 4c(t) > 0$. Meanwhile, $x(0) = 0$ and $c(0) = 0$. These facts mean that all the conditions of Lemma 2.1 hold. Thus, by Lemma 2.1, we have

$$\|L^{-1}(\Delta L)\|_F = x(1) \leq \frac{1}{2} \left(\sqrt{2} - \sqrt{2 - 4c(1)} \right) = \frac{1}{\sqrt{2}} \left(1 - \sqrt{1 - 2\|L^{-1}\|_2^2 \|\Delta K\|_F} \right). \quad (3.8)$$

Now we introduce a scaling matrix $D_{m+n} \in \mathbb{D}_{m+n}$ into the expression (3.7) with $t = 1$, which can be used to improve the bounds. Right-multiplying (3.7) with $t = 1$ by D_{m+n} and using (2.4) yields

$$J_{m+n}(\Delta L)^T L^{-T} D_{m+n} = \text{up} \left(L^{-1}(\Delta K) L^{-T} D_{m+n} - L^{-1}(\Delta L) J_{m+n}(\Delta L)^T L^{-T} D_{m+n} \right). \quad (3.9)$$

Taking the Frobenius norm on (3.9) and noting (2.2) and (2.1), we get

$$\|(\Delta L)^T L^{-T} D_{m+n}\|_F \leq \|L^{-1}\|_2 \|L^{-T} D_{m+n}\|_2 \|\Delta K\|_F + \|L^{-1} \Delta L\|_F \|(\Delta L)^T L^{-T} D_{m+n}\|_F,$$

which combined with (3.8) gives

$$\|(\Delta L)^T L^{-T} D_{m+n}\|_F \leq \frac{\|L^{-1}\|_2 \|L^{-T} D_{m+n}\|_2 \|\Delta K\|_F}{1 - \|L^{-1} \Delta L\|_F} \leq \frac{\sqrt{2} \|L^{-1}\|_2 \|L^{-T} D_{m+n}\|_2 \|\Delta K\|_F}{\sqrt{2} - 1 + \sqrt{1 - 2\|L^{-1}\|_2^2 \|\Delta K\|_F}}. \quad (3.10)$$

Note that $(\Delta L)^T = (\Delta L)^T L^{-T} D_{m+n} (L^{-T} D_{m+n})^{-1}$. Then, from (3.10), we have

$$\|\Delta L\|_F = \|(\Delta L)^T\|_F \leq \frac{\sqrt{2} \|L^{-1}\|_2 \|L^{-T} D_{m+n}\|_2 \|(L^{-T} D_{m+n})^{-1}\|_2 \|\Delta K\|_F}{\sqrt{2} - 1 + \sqrt{1 - 2\|L^{-1}\|_2^2 \|\Delta K\|_F}}.$$

Since

$$\|L^{-T} D_{m+n}\|_2 \|(L^{-T} D_{m+n})^{-1}\|_2 = \|(D_{m+n}^{-1} L^T)^{-1}\|_2 \|D_{m+n}^{-1} L^T\|_2 = \kappa_2(D_{m+n}^{-1} L^T) = \kappa_2(L D_{m+n}^{-1})$$

and $D_{m+n} \in \mathbb{D}_{m+n}$ is arbitrary, we have the bound (3.3). The bound (3.4) follows from (3.3). \square

Now we give some remarks on this theorem, which are analogous to those in [7] on [7, Theorem 3.1].

Remark 3.1. Taking the infimum of the expression below (21) in [2] over the set \mathbb{D}_{m+n} and adding the higher-order term, we can derive the following first-order perturbation bound:

$$\|\Delta L\|_F \leq \|L^{-1}\|_2 \left[\inf_{D_{m+n} \in \mathbb{D}_{m+n}} \kappa_2(L D_{m+n}^{-1}) \right] \|\Delta K\|_F + O(\|\Delta K\|_F^2). \quad (3.11)$$

It is easy to find that the difference between the first-order bound (3.11) and the rigorous bound (3.4) is a factor of $2 + \sqrt{2}$.

Remark 3.2. In [3, Theorem 2.3], the author obtained the following rigorous perturbation bound by the classic matrix equation approach:

$$\|\Delta L\|_F \leq \frac{\sqrt{2} \|L^{-1}\|_2 \kappa_2(L) \|\Delta K\|_F}{1 + \sqrt{1 - 2\|L^{-1}\|_F^2 \|\Delta K\|_F}} \quad (3.12)$$

under the condition

$$\|L^{-1}\|_F^2 \|\Delta K\|_F < \frac{1}{2}.$$

The bound (3.12) is a little larger than

$$\|\Delta L\|_F \leq \frac{\sqrt{2} \|L^{-1}\|_2 \kappa_2(L)}{1 + \sqrt{1 - 2 \|L^{-1}\|_2^2 \|\Delta K\|_F}} \|\Delta K\|_F. \quad (3.13)$$

Setting $D_{m+n} = I_{m+n}$ in (3.3) gives

$$\|\Delta L\|_F \leq \frac{\sqrt{2} \|L^{-1}\|_2 \kappa_2(L)}{\sqrt{2} - 1 + \sqrt{1 - 2 \|L^{-1}\|_2^2 \|\Delta K\|_F}} \|\Delta K\|_F. \quad (3.14)$$

In comparison, we can find that the bound (3.14) is at most $\sqrt{2} + 1$ times as large as the bound (3.13). However, $\inf_{D_{m+n} \in \mathbb{D}_{m+n}} \kappa_2(LD_{m+n}^{-1})$ can be arbitrarily smaller than $\kappa_2(L)$ when L has bad column scaling.

For example, let $L = \begin{bmatrix} 1/\gamma & 0 \\ 1 & 1 \end{bmatrix}$ with $0 < \gamma \ll 1$. Then $\kappa_2(LD_{m+n}^{-1}) = \Theta(1)$ with $D_{m+n} = \text{diag}(1/\gamma, 1)$, and $\kappa_2(L) = \Theta(1/\gamma)$. Therefore, the bound (3.3) can be much sharper than (3.13) and hence (3.12).

Remark 3.3. The following rigorous perturbation bound given in [6, Theorem 3.1] was derived by the matrix-vector equation approach:

$$\|\Delta L\|_F \leq 2 \left\| \widehat{W}_{JL^T}^{-1} \right\|_2 \|\Delta K\|_F, \quad (3.15)$$

where \widehat{W}_{JL^T} is a $\frac{(m+n)(m+n+1)}{2} \times \frac{(m+n)(m+n+1)}{2}$ lower triangular matrix defined by the elements of JL^T , under the condition

$$\left\| \widehat{W}_{JL^T}^{-1} \right\|_2 \left\| \widehat{W}_{JL^T}^{-1} \text{duvec}(\Delta K) \right\|_2 < \frac{1}{4},$$

which can be as bad as

$$\left\| \widehat{W}_{JL^T}^{-1} \right\|_2^2 \|\Delta K\|_F < \frac{1}{4}. \quad (3.16)$$

Please see [6, 9] for the specific structure of the matrix \widehat{W}_{JL^T} and the definition of "duvec."

Numerical experiments indicated that the bound (3.15) is a little tighter than (3.4), however, it is not much tighter than (3.4). But the condition (3.16) can be much stronger than (3.1). For example, let $L = \begin{bmatrix} 1 & 0 \\ \gamma & 1 \end{bmatrix}$ with $\gamma \gg 1$. Then $\left\| \widehat{W}_{JL^T}^{-1} \right\|_2^2 \approx \Theta(\gamma^4)$ and $\|L^{-1}\|_2^2 = \Theta(\gamma^2)$. Moreover, it is more expensive to compute the bound (3.15) than that of (3.4).

Remark 3.4. The following rigorous perturbation bound was derived by the refined matrix equation approach [6, Theorem 3.2]:

$$\begin{aligned} \|\Delta L\|_F &\leq \frac{2 \|L\|_2 \kappa_2(L) \kappa_2(LD_{m+n}^{-1}) \frac{\|\Delta K\|_F}{\|K\|_2}}{1 + \sqrt{1 - 4 \kappa_2(L) \|L\|_2 \|D_{m+n} L^{-1}\|_2 \|D_{m+n}^{-1}\|_2 \frac{\|\Delta K\|_F}{\|K\|_2}}} \\ &\leq 2 \|L\|_2 \kappa_2(L) \kappa_2(LD_{m+n}^{-1}) \frac{\|\Delta K\|_F}{\|K\|_2} \end{aligned} \quad (3.17)$$

under the condition

$$\kappa_2(L) \|L\|_2 \|D_{m+n} L^{-1}\|_2 \|D_{m+n}^{-1}\|_2 \frac{\|\Delta K\|_F}{\|K\|_2} < \frac{1}{4}. \quad (3.18)$$

Minimizing $\kappa_2(LD_{m+n}^{-1})$ over the set \mathbb{D}_{m+n} , we can make the bound (3.17) similar to the new bound (3.4). However, noting the facts $\|K\|_2 \leq \|L\|_2^2$ and $\|D_{m+n}L^{-1}\|_2 \geq \|L^{-1}\|_2 / \|D_{m+n}^{-1}\|_2$, we have

$$\kappa_2(L) \|L\|_2 \|D_{m+n}L^{-1}\|_2 \|D_{m+n}^{-1}\|_2 \frac{\|\Delta K\|_F}{\|K\|_2} \geq \|L^{-1}\|_2^2 \|\Delta K\|_F.$$

Therefore, the condition (3.18) is not only complicated but also more constraining than (3.1). Especially, when we minimize the bound (3.17) over the set \mathbb{D}_{m+n} , the best choice of $D \in \mathbb{D}_{m+n}$ may make the condition (3.18) worse.

From the above discussions, we can find that the bounds given in Theorem 3.1 have more advantages compared with the existing ones.

4 Rigorous perturbation bounds with componentwise perturbation

According to Algorithm 1 for computing the generalized Cholesky factorization given in [1], [1, Eqns. (8)–(10)], and [10, Lemma 8.4 and Theorem 10.3], we have that if

$$K + \Delta K = \tilde{L}J_{m+n}\tilde{L}^T, \quad \tilde{L} = \begin{pmatrix} \tilde{L}_{11} & 0 \\ \tilde{L}_{21} & \tilde{L}_{22} \end{pmatrix}, \quad \Delta K = \begin{pmatrix} \Delta A & (\Delta B)^T \\ \Delta B & \Delta C \end{pmatrix},$$

then

$$|\Delta A| \leq \gamma_{3m+1} |\tilde{L}_{11}| |\tilde{L}_{11}^T|, \quad |\Delta B| \leq \gamma_{3m+1} |\tilde{L}_{21}| |\tilde{L}_{11}^T|, \quad |\Delta C| \leq \gamma_{3n+1} |\tilde{L}_{22}| |\tilde{L}_{22}^T|,$$

where $\gamma_i = iu/(1-iu)$ and u is the unit roundoff. Thus, the computed generalized Cholesky factor \tilde{L} satisfies

$$K + \Delta K = \tilde{L}J_{m+n}\tilde{L}^T, \quad |\Delta K| \leq \varepsilon |\tilde{L}^T| |\tilde{L}|, \quad (4.1)$$

where $\varepsilon = \min\{\gamma_{3m+1}, \gamma_{3n+1}\}$.

In the following, we consider the rigorous perturbation bounds for the generalized Cholesky factorization with the perturbation ΔK having the same form as in (4.1).

Theorem 4.1. *Let $\Delta K \in \mathbb{R}^{(m+n) \times (m+n)}$ be a symmetric perturbation in $K \in \mathbb{R}^{(m+n) \times (m+n)}$ which has the same form as that in (1.1), however, does not necessarily have the generalized Cholesky factorization, and $K + \Delta K$ have the generalized Cholesky factorization (4.1). If*

$$\text{cond}_F(\tilde{L}) \text{cond}_F(\tilde{L}^{-T}) \varepsilon < \frac{1}{2}, \quad (4.2)$$

then K has the generalized Cholesky factorization $K = LJ_{m+n}L^T$, where $L = \tilde{L} - \Delta L$. Moreover,

$$\|\Delta L\|_2 \leq \frac{\sqrt{2} \left(\inf_{D_{m+n} \in \mathbb{D}_{m+n}} \left\| \tilde{L}D_{m+n}^{-1} \right\|_2 \left\| D_{m+n} \tilde{L}^{-1} \right\|_2 \right) \text{cond}_F(\tilde{L}) \varepsilon}{\sqrt{2} - 1 + \sqrt{1 - 2 \text{cond}_F(\tilde{L}) \text{cond}_F(\tilde{L}^{-T}) \varepsilon}} \quad (4.3)$$

$$\leq (2 + \sqrt{2}) \left(\inf_{D_{m+n} \in \mathbb{D}_{m+n}} \left\| \tilde{L}D_{m+n}^{-1} \right\|_2 \left\| D_{m+n} \tilde{L}^{-1} \right\|_2 \right) \text{cond}_F(\tilde{L}) \varepsilon. \quad (4.4)$$

Proof. From (2.1) and (4.1), for any $t \in [0, 1]$, we have

$$\left\| \tilde{L}^{-1} t(\Delta K) \tilde{L}^{-T} \right\|_F \leq \left\| \tilde{L}^{-1} \|\tilde{L}\| \tilde{L}^T \|\tilde{L}^{-T}\| \right\|_F \varepsilon \leq \text{cond}_F(\tilde{L}) \text{cond}_F(\tilde{L}^{-T}) \varepsilon < \frac{1}{2}.$$

Thus, considering the proof of Theorem 3.1, for any $t \in [0, 1]$, $(K + \Delta K) - t(\Delta K)$ has the following generalized Cholesky factorization

$$(K + \Delta K) - t(\Delta K) = \tilde{L}(t)J_{m+n}\tilde{L}^T(t) = (\tilde{L} - (\Delta L(t)))J_{m+n}(\tilde{L} - (\Delta L(t)))^T, \quad (4.5)$$

where $\tilde{L}(t)$ is lower triangular having the same structure as that of L in (1.2) and $\Delta L(t) = \tilde{L} - \tilde{L}(t)$ with $\Delta L(0) = 0$. Setting $\Delta L(1) = \Delta L$ in (4.5) leads to the generalized Cholesky factorization of K as in (1.2).

Next, we prove (4.3) and (4.4). The proof is similar to the one for the bounds (3.3) and (3.4). Considering (4.1), from (4.5), we have

$$J_{m+n}(\Delta L(t))^T \tilde{L}^{-T} = \text{up} \left[t\tilde{L}^{-1}(\Delta K)\tilde{L}^{-T} + \tilde{L}^{-1}(\Delta L(t))J_{m+n}(\Delta L(t))^T \tilde{L}^{-T} \right]. \quad (4.6)$$

Taking the Frobenius norm on (4.6), and using (2.3), (2.1) and the bound of $|\Delta K|$ in (4.1) gives

$$\left\| \tilde{L}^{-1}(\Delta L(t)) \right\|_F = \left\| J_{m+n}(\Delta L(t))^T \tilde{L}^{-T} \right\|_F \leq \frac{1}{\sqrt{2}} \left[\text{cond}_F(\tilde{L})\text{cond}_F(\tilde{L}^{-T})\varepsilon t + \left\| \tilde{L}^{-1}(\Delta L(t)) \right\|_F^2 \right].$$

That is,

$$x^2(t) - \sqrt{2}x(t) + c(t) \geq 0,$$

where $x(t) = \left\| \tilde{L}^{-1}(\Delta L(t)) \right\|_F$ and $c(t) = \text{cond}_F(\tilde{L})\text{cond}_F(\tilde{L}^{-T})\varepsilon t$. Considering (4.2), we can check that the conditions of Lemma 2.1 hold. Thus,

$$\left\| \tilde{L}^{-1}(\Delta L) \right\|_F \leq \frac{1}{\sqrt{2}} \left(1 - \sqrt{1 - 2\text{cond}_F(\tilde{L})\text{cond}_F(\tilde{L}^{-T})\varepsilon} \right). \quad (4.7)$$

Taking $t = 1$ in (4.6) and considering (2.4), for any $D_{m+n} \in \mathbb{D}_{m+n}$, we have

$$J_{m+n}(\Delta L)^T \tilde{L}^{-T} D_{m+n} = \text{up} \left(\tilde{L}^{-1}(\Delta K)\tilde{L}^{-T} D_{m+n} + \tilde{L}^{-1}(\Delta L)J_{m+n}(\Delta L)^T \tilde{L}^{-T} D_{m+n} \right). \quad (4.8)$$

Taking the Frobenius norm on (4.8) and using (2.2) and (2.1) yields

$$\left\| D_{m+n}\tilde{L}^{-1}(\Delta L) \right\|_F \leq \left\| \tilde{L}^{-1}(\Delta K)\tilde{L}^{-T} D_{m+n} \right\|_F + \left\| \tilde{L}^{-1}(\Delta L) \right\|_F \left\| D_{m+n}\tilde{L}^{-1}(\Delta L) \right\|_F.$$

Further, substituting the bound of $|\Delta K|$ in (4.1) into the above inequality leads to

$$\begin{aligned} \left\| D_{m+n}\tilde{L}^{-1}(\Delta L) \right\|_F &\leq \left\| \tilde{L}^{-1}|\tilde{L}|\tilde{L}^T|\tilde{L}^{-T} D_{m+n} \right\|_F \varepsilon + \left\| \tilde{L}^{-1}(\Delta L) \right\|_F \left\| D_{m+n}\tilde{L}^{-1}(\Delta L) \right\|_F \\ &\leq \left\| D_{m+n}|\tilde{L}^{-1}||\tilde{L}| \right\|_2 \text{cond}_F(\tilde{L})\varepsilon + \left\| \tilde{L}^{-1}(\Delta L) \right\|_F \left\| D_{m+n}\tilde{L}^{-1}(\Delta L) \right\|_F. \end{aligned}$$

Noting (4.7), we obtain

$$\left\| D_{m+n}\tilde{L}^{-1}(\Delta L) \right\|_F \leq \frac{\sqrt{2} \left\| D_{m+n}|\tilde{L}^{-1}||\tilde{L}| \right\|_2 \text{cond}_F(\tilde{L})\varepsilon}{\sqrt{2} - 1 + \sqrt{1 - 2\text{cond}_F(\tilde{L})\text{cond}_F(\tilde{L}^{-T})\varepsilon}},$$

which combined with the fact

$$\|\Delta L\|_F \leq \left\| \tilde{L}D_{m+n}^{-1} \right\|_2 \left\| D_{m+n}\tilde{L}^{-1}(\Delta L) \right\|_F$$

implies the bound (4.3) and then (4.4). \square

Remark 4.1. The following first-order perturbation bound can be derived from (4.3),

$$\|\Delta L\|_2 \leq \left(\inf_{D_{m+n} \in \mathbb{D}_{m+n}} \left\| \tilde{L}D_{m+n}^{-1} \right\|_2 \left\| D_{m+n}|\tilde{L}^{-1}||\tilde{L}| \right\|_2 \right) \text{cond}_F(\tilde{L})\varepsilon + O(\varepsilon^2). \quad (4.9)$$

In addition, it is worthy pointing out that the perturbation bounds obtained in this section are unlike the ones in Section 3. These bounds involve the generalized Cholesky factor of $K + \Delta K$ but not the one of K . This is because the bound of $|\Delta K|$ in (4.1) involves the generalized Cholesky factor of $K + \Delta K$.

5 Concluding remarks

In this paper, some new rigorous perturbation bounds for the generalized Cholesky factorization with normwise or componentwise perturbations in the given matrix are obtained. These bounds not only have simple and moderate conditions but also can be much smaller than some existing ones. To estimate these bounds efficiently, the suitable scaling matrix D is needed. In [9], the author provided some methods, which are also applicable to the bounds in this paper. Please refer to [9] for detail on these methods.

References

- [1] J.-X. Zhao, The generalized Cholesky factorization method for saddle point problems, *Appl. Math. Comput.* 92 (1998) 49–58.
- [2] W.-G. Wang, J.-X. Zhao, Perturbation analysis for the generalized Cholesky factorization, *Appl. Math. Comput.* 147 (2004) 601–606.
- [3] W.-G. Wang, Perturbation bounds for the generalized Cholesky factorization, *Nanjing Univ. J. Math. Biquarterly.* 20 (2003) 113–119.
- [4] X.-M. Fang, The generalized Cholesky factorization and perturbation of the real symmetric matrices, *J. Zhaoqing Univ.* 29 (5) (2008) 14–16.
- [5] X.-M. Fang, New perturbation analysis for the generalized Cholesky factorization, *J. Zhanjiang Norm. College* 30 (3) (2009) 16–20.
- [6] H.-Y. Li, H. Yang, H. Shao, A note on the perturbation analysis for the generalized Cholesky factorization, *Appl. Math. Comput.* 215 (2010) 4022–4027.
- [7] X.-W. Chang, D. Stehlé, Rigorous perturbation bounds of some matrix factorizations, *SIAM J. Matrix Anal. Appl.* 31 (2010) 2841–2859.
- [8] G.W. Stewart, J.-G. Sun, *Matrix Perturbation Theory*, Academic Press, Boston, 1990.
- [9] X.-W. Chang, *Perturbation Analysis of Some Matrix Factorizations*, Ph.D. Thesis, School of Computer Science, McGill University, Montreal, 1997.
- [10] N. J. Higham, *Accuracy and stability of numerical algorithms*, SIAM Publications, Philadelphia, PA, 2002.